

Aplicações Informáticas na Biomedicina

3.ª Aula Prática Laboratorial

Mestrado Integrado em Engenharia Informática

Ano Letivo 2019/2020

Marisa Esteves

9 de Outubro de 2019



Universidade do Minho

Plano de Aula

1. Finalização da resolução da 2.^a ficha prática laboratorial pelos alunos em grupo;
2. Correção da 2.^a ficha prática laboratorial com os alunos;
3. Contextualização sobre os processos de ETL (*Extract, Transform, Load*) e de data warehousing.

Processo ETL

Definição

O processo ETL (*Extract, Transform, Load*) é um conjunto de processos que inclui a extração de dados de fontes de informação internas e externas, podendo estar em diferentes formatos, a transformação dos dados de acordo com as necessidades da organização e, finalmente, o carregamento dos mesmos numa estrutura de dados, como por exemplo um data mart ou um data warehouse.

Processo ETL

Definição

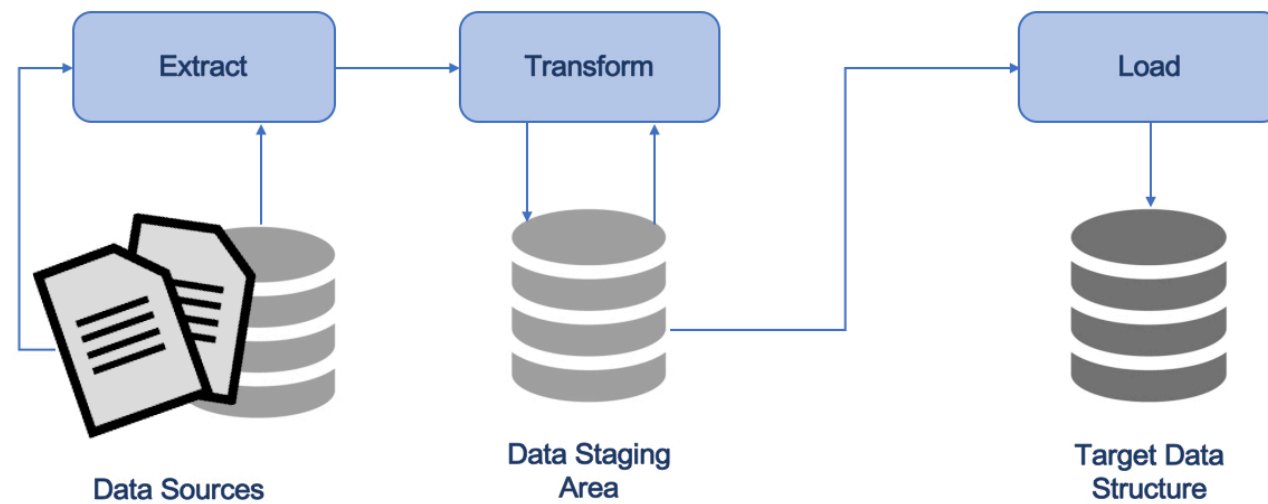


Figura 1 – Esquema do processo ETL.

Processo ETL

Porquê?

Os dados estão espalhados
por diferentes localizações

Os dados estão
armazenados em diferentes
tipos de formato

O volume de dados continua
a aumentar

Os dados podem estar
estruturados, semi-
estruturados ou não
estruturados

Data Warehousing

Definição

O processo de data warehousing enfatiza à recolha de dados de diversas fontes através do processo ETL (*Extract, Transform, Load*), correspondendo à construção de data warehouses e/ou data marts, para aceder e analisar a informação de forma útil. Os dados extraídos são processados, formatados e consolidados numa estrutura de dados única para facilitar essencialmente a análise de dados.

Data Warehousing

Definição

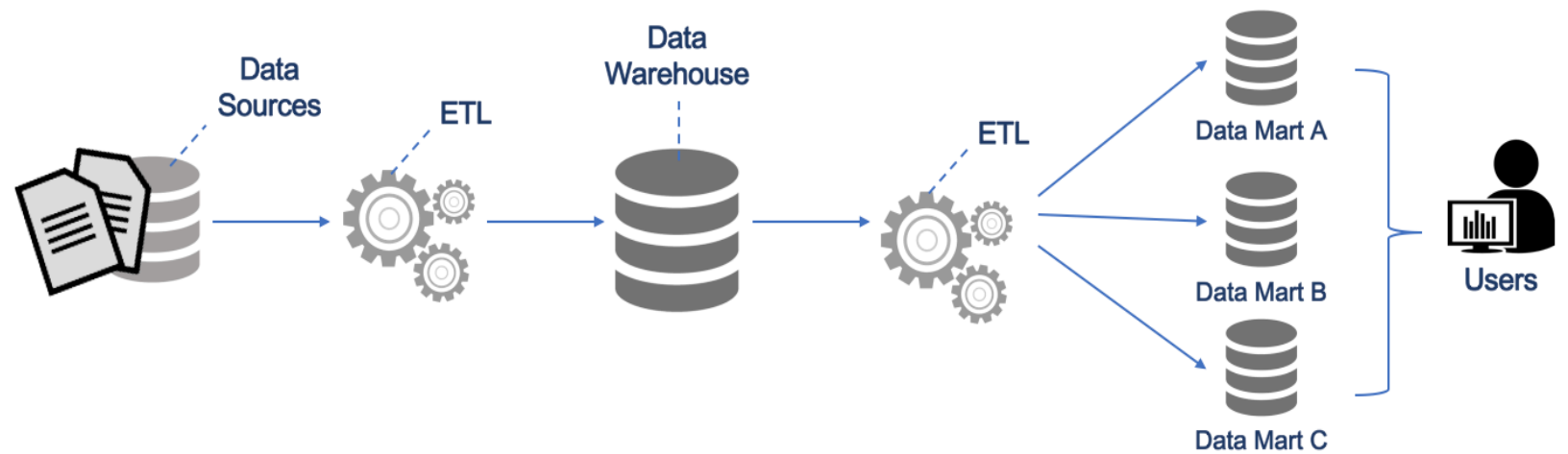


Figura 2 – Esquema do processo de data warehousing.

Data Warehousing

*Data Warehouse vs. Data
Mart*

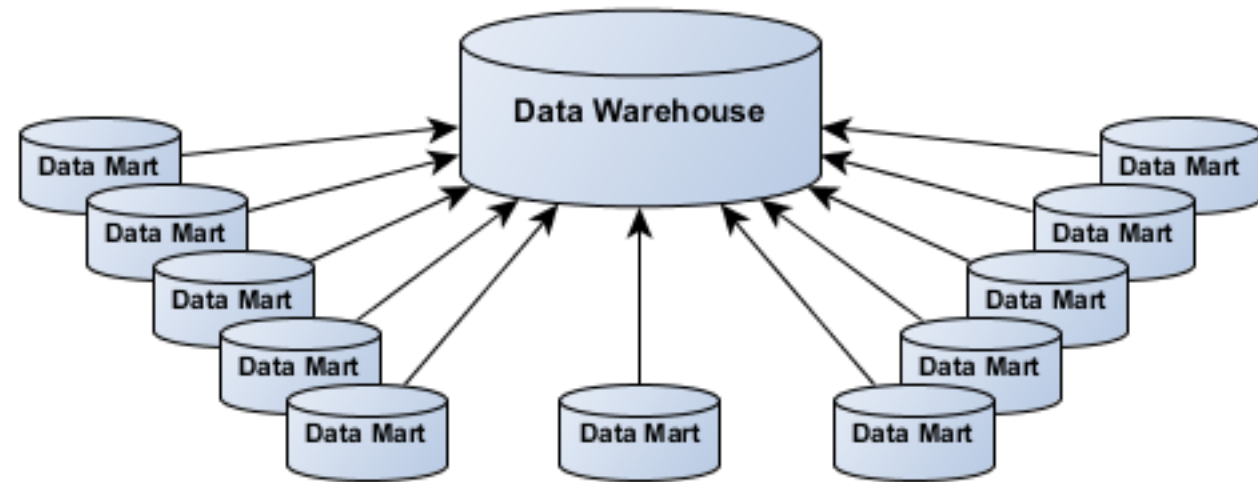


Figura 3 – Data warehouse vs. Data marts.

Data Warehousing

*Modelo Dimensional –
Esquema em Estrela vs.
Esquema em Floco de Neve*

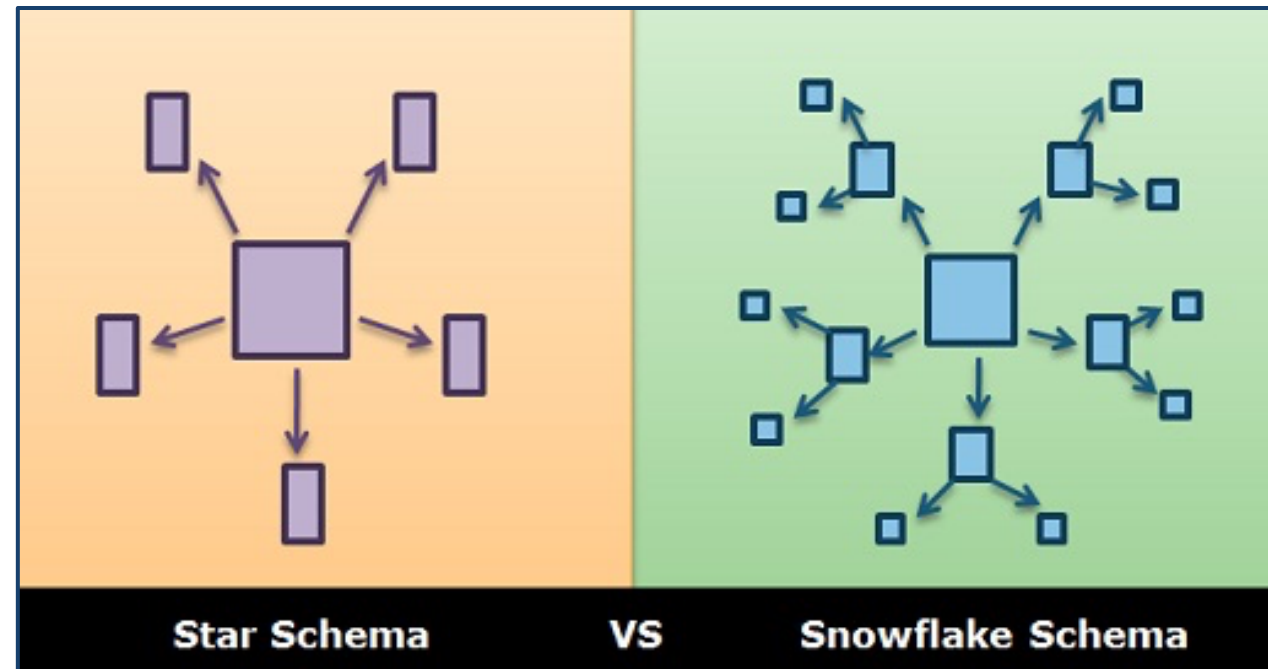


Figura 4 – Esquema em Estrela vs. Esquema em Floco de Neve.

Data Warehousing

*Modelo Dimensional –
Esquema em Constelação
de Factos*

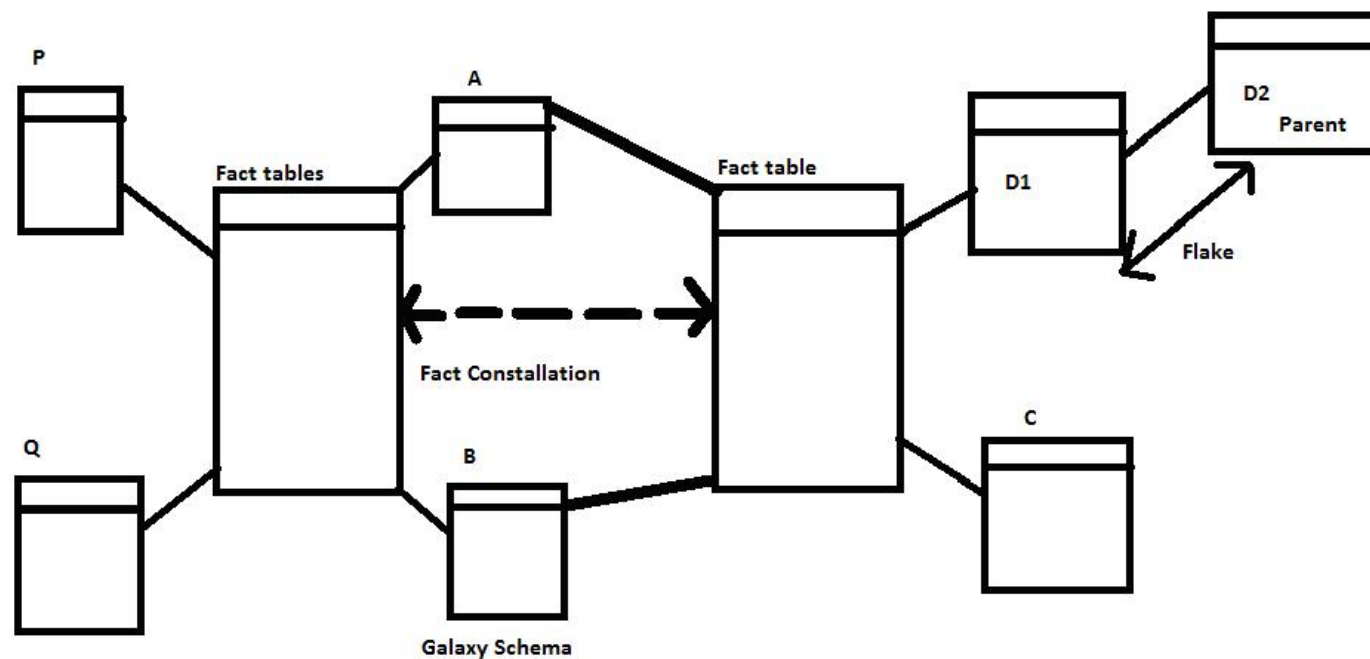


Figura 5 – Esquema em Constelação de Factos.

OLTP vs. OLAP

Definição

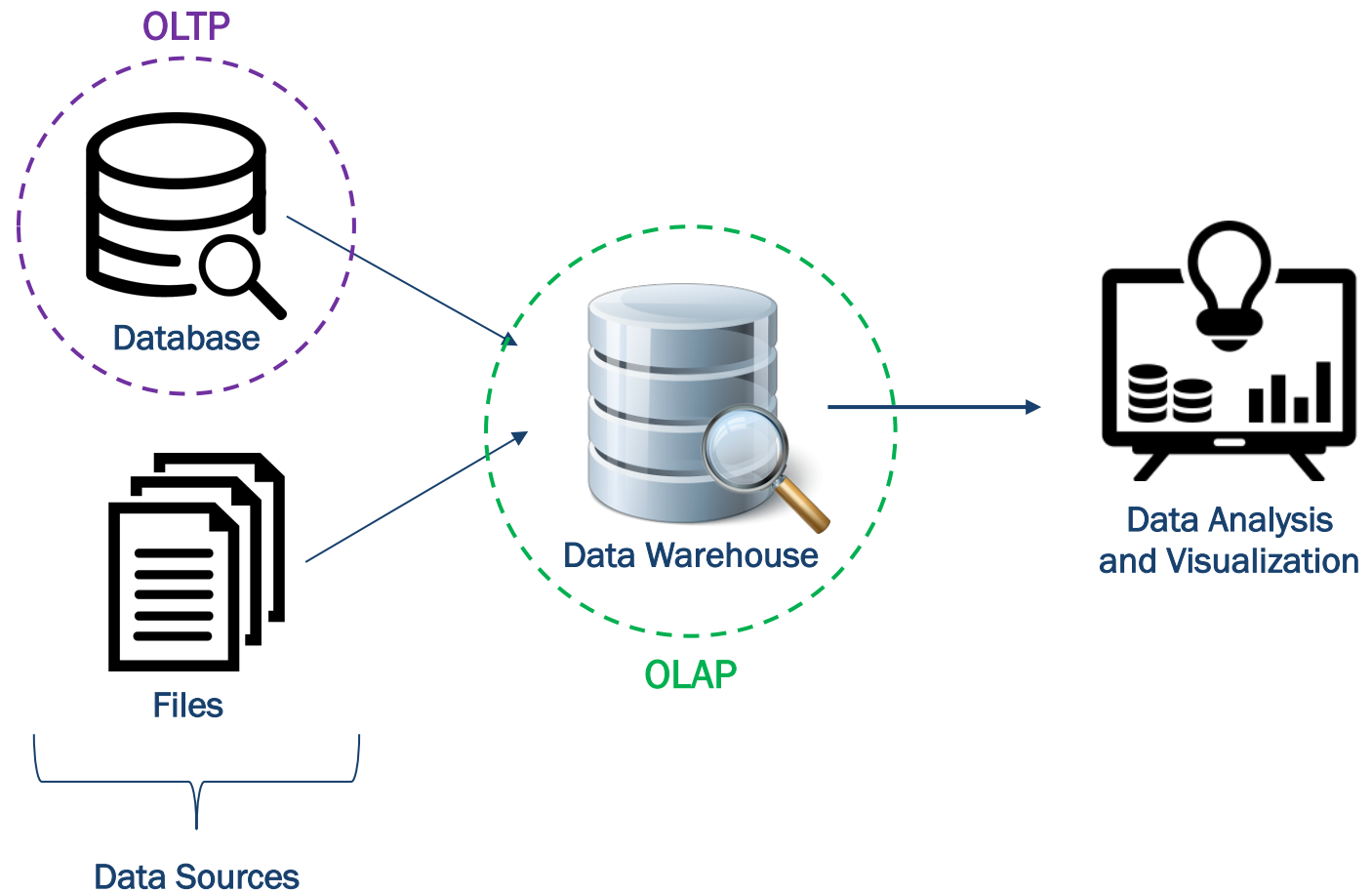


Figura 6 – OLTP (*Online Transaction Processing*) vs. OLAP (*Online Analytical Processing*).

OLTP vs. OLAP

Definição

Relational Database (OLTP)	Analytical Data Warehouse (OLAP)
Contains current data	Contains historical data
Useful in running the business	Useful in analysing the business
Based on Entity Relationship Model	Based on Star, Snowflake or Fact Constellation Schema
Provides primitive and highly detailed data	Provides summarized and consolidated data
Used for writing into the database	Used for reading data from the data warehouse
Database size ranges from 100 MB to 1 GB	Data warehouse ranges from 100 GB to 1 TB
Fast and it provides high performance	Highly flexible but it is not fast
Number of records accessed is in tens	Number of records accessed is in millions
Example: all bank transactions made by a customer	Example: bank transactions made by a customer at a particular time

Figura 7 – Diferenças entre OLTP e OLAP.

MySQL

**INSERT INTO
SELECT FROM**

Permite copiar dados de uma tabela e os inserir noutra tabela. No entanto, este comando SQL requer que os tipos de dados na tabela de origem (table1) e na tabela destino (table2) sejam iguais.

- **INSERT INTO** *table2* (*column1*, *column2*, *column3*, ...) **SELECT** *column1*, *column2*, *column3*, ... **FROM** *table1* **WHERE** *condition*

MySQL

Cursores

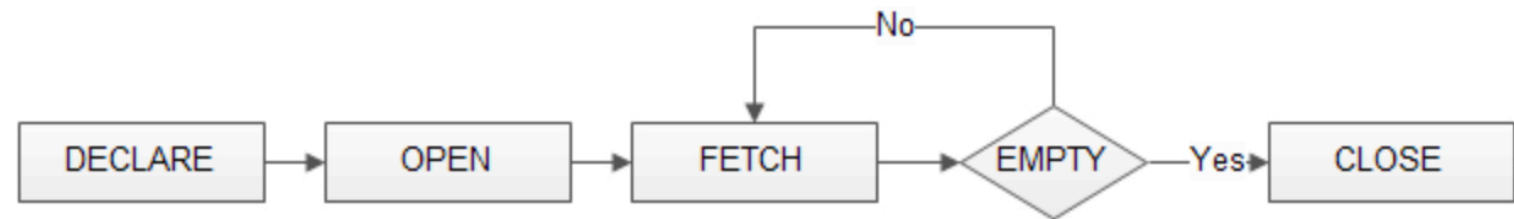


Figura 8 – Modo de funcionamento de cursores em MySQL.

MySQL

Cursores

```
1 CREATE PROCEDURE curdemo()  
2 BEGIN  
3     DECLARE done INT DEFAULT FALSE;  
4     DECLARE a CHAR(16);  
5     DECLARE b, c INT;  
6     DECLARE cur1 CURSOR FOR SELECT id,data FROM test.t1;  
7     DECLARE cur2 CURSOR FOR SELECT i FROM test.t2;  
8     DECLARE CONTINUE HANDLER FOR NOT FOUND SET done = TRUE;  
9  
10    OPEN cur1;  
11    OPEN cur2;  
12  
13    read_loop: LOOP  
14        FETCH cur1 INTO a, b;  
15        FETCH cur2 INTO c;  
16        IF done THEN  
17            LEAVE read_loop;  
18        END IF;  
19        IF b < c THEN  
20            INSERT INTO test.t3 VALUES (a,b);  
21        ELSE  
22            INSERT INTO test.t3 VALUES (a,c);  
23        END IF;  
24    END LOOP;  
25  
26    CLOSE cur1;  
27    CLOSE cur2;  
28 END;
```

Figura 9 – Exemplo de um procedimento com cursores em MySQL.

Resolução da 2.ª Ficha Prática Laboratorial

1 Registo da Realização de Consultas numa Clínica

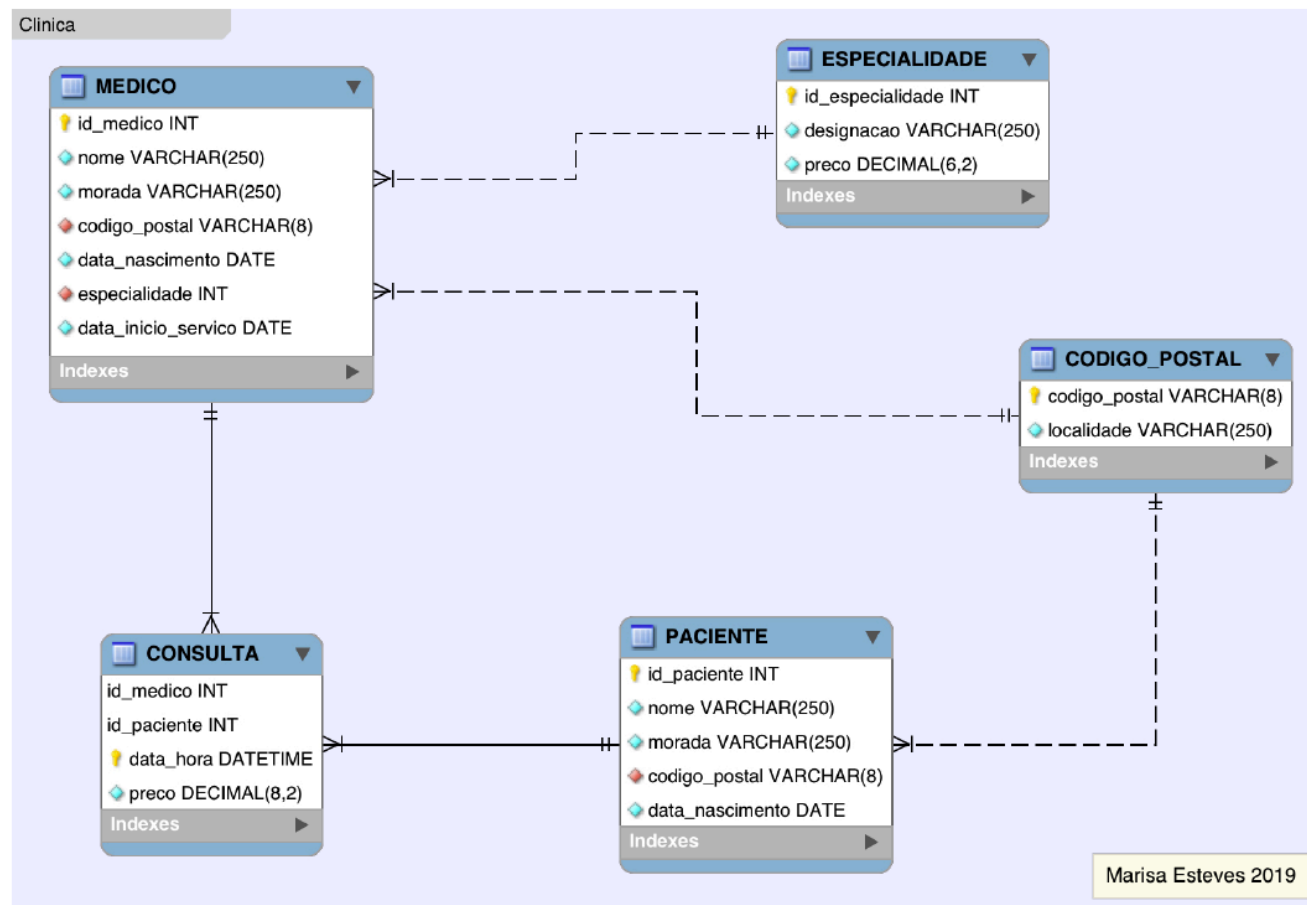


Figura 1: Esquema lógico da base de dados de uma determinada clínica.

Resolução da 2.ª Ficha Prática Laboratorial

Na Figura 1 é apresentado o esquema lógico, nomeadamente modelo relacional, da base de dados de uma determinada clínica.

A base de dados que o esquema representa está integrada no sistema operacional principal da clínica e suporta todos os processos relacionados com o registo e a faturação das consultas realizadas. É de referir que o valor pago pelo paciente em cada consulta pode ser superior ao valor de referência para a especialidade do médico (preço), uma vez que podem ser pagos procedimentos efetuados durante a realização da consulta.

Com base no caso apresentado, pretende-se que:

1. Instale o sistema de gestão de bases de dados relacionais MySQL:
 - (a) Sistema operativo Windows (MySQL Installer): <https://dev.mysql.com/downloads/installer/>;
 - (b) Sistema operativo macOS (MySQL Community Server & MySQL Workbench): <https://dev.mysql.com/downloads/>.
2. Utilizando o MySQL Workbench, e o ficheiro “Ficha2_ModeloLogico.mwb” enviado juntamente com esta ficha prática laboratorial, faça a geração do respetivo esquema físico para a base de dados em questão (Database > Forward Engineer).

Resolução da 2.ª Ficha Prática Laboratorial

3. Povoie a base de dados criada, utilizando o ficheiro “Ficha2_PovoarTabelas.sql” disponibilizado juntamente com esta ficha prática laboratorial. O ficheiro disponibiliza igualmente a função “idade” que será muito útil para responder a alguns dos pontos da questão 4.
4. Utilizando SQL, desenvolva as *queries* necessárias para responder às seguintes questões:
 - (a) Qual é o nome dos médicos com mais de 10 anos de serviço?
 - (b) Qual é o nome de todos os médicos e a respetiva especialidade que cada um exerce?
 - (c) Qual é o nome e a idade dos médicos com mais de 40 anos de idade da especialidade de Clínica Geral?
 - (d) Qual é o nome e os anos de serviço dos médicos com mais de 50 anos de idade que deram consultas a partir das 12h a pacientes com menos de 20 anos de idade?
 - (e) Qual é o nome dos médicos com mais de 30 anos de idade ou menos de 5 anos de serviço?
 - (f) Qual é o nome e a idade dos pacientes que já foram consultados por todos os médicos?
 - (g) Qual é o nome dos médicos que nunca consultaram pacientes residentes em Braga?
 - (h) Apresente a média dos anos de serviço dos médicos para cada uma das especialidades. Devem ser apresentadas todas as especialidades, incluindo as que não tenham médicos associados.
 - (i) Apresente o número de consultas que estão registadas por cada um dos médicos. Devem ser apresentados todos os médicos, incluindo os que nunca tenham dado consultas.
 - (j) Apresente para cada médico o valor total faturado em 2017. Devem ser apresentados todos os médicos, incluindo os que nunca tenham dado consultas.
 - (k) Apresente o número de médicos para cada uma das especialidades. Devem ser apresentadas todas as especialidades, incluindo as que não tenham médicos associados.
5. Adicione um atributo denominado “total_faturado” na tabela *MEDICO* para acumular os valores faturados por cada um dos médicos nas suas consultas. Numa primeira etapa, pretende-se que este atributo seja carregado recorrendo aos dados já existentes na base de dados. Seguidamente, efetue igualmente as operações necessárias para que o referido atributo se mantenha sempre atualizado.